The Compact Muon Solenoid Experiment

Conference Report

Mailing address: CMS CERN, CH-1211 GENEVA 23, Switzerland

28 June 2017 (v5, 05 July 2017)

Track Finding for the Level-1 Trigger of the CMS Experiment

Tom James on behalf of the TMTT group

Abstract

A new tracking system is under development for the CMS experiment at the High Luminosity LHC (HL-LHC), located at CERN. It includes a silicon tracker that will correlate clusters in two closely spaced sensor layers, for the rejection of hits from low transverse momentum tracks. This will allow tracker data to be read out to the Level-1 trigger at 40 MHz. The Level-1 track-finder must be able to identify tracks with transverse momentum above 2–3 GeV/ c within latency constraints. A concept for an FPGA-based track finder using a fully time-multiplexed architecture is presented, where track candidates are identified using a Hough Transform, and then refined with a Kalman Filter. Both steps are fully implemented in FPGA firmware. A hardware system built from MP7 MicroTCA processing cards has been assembled, which demonstrates a realistic slice of the track finder in order to help gauge the performance and requirements for a final system.

Presented at *International Conference on Technology and Instrumentation in Particle Physics TIPP2017*

Track Finding for the Level-1 Trigger of the CMS Experiment

Tom James tom.james@cern.ch
on behalf of the time-multiplexed track-trigger group
and the CMS collaboration

Imperial College London

Abstract. A new tracking system is under development for the CMS experiment at the High Luminosity LHC (HL-LHC), located at CERN. It includes a silicon tracker that will correlate clusters in two closely spaced sensor layers, for the rejection of hits from low transverse momentum tracks. This will allow tracker data to be read out to the Level-1 trigger at 40 MHz. The Level-1 track-finder must be able to identify tracks with transverse momentum above 2–3 GeV/ c within latency constraints. A concept for an FPGA-based track finder using a fully time-multiplexed architecture is presented, where track candidates are identified using a Hough Transform, and then refined with a Kalman Filter. Both steps are fully implemented in FPGA firmware. A hardware system built from MP7 MicroTCA processing cards has been assembled, which demonstrates a realistic slice of the track finder in order to help gauge the performance and requirements for a final system.

1 The High-Luminosity LHC and the new CMS tracker

The Compact Muon Solenoid (CMS) detector [1] at the CERN Large Hadron Collider (LHC) [2] is an all-purpose detector, designed to investigate a wide range of physics. Beginning in 2026, the LHC will operate at an upgraded luminosity of $5\text{--}7.5 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$, corresponding to 140–200 simultaneous collisions (pileup) at 40 MHz [3]. This will enable the collection of 3000 fb^{-1} of data by 2035.

Due to accumulated radiation damage, it will be necessary to completely replace the CMS outer silicon tracker during the long LHC shutdown preceding this upgrade. Studies have shown that the use of tracker information in the level-1 (L1) trigger will be required to keep the accept rate below the target 750 kHz, without significant degradation in physics sensitivity. A new silicon tracker has therefore been designed to allow the read out of a sub-set of its data at 40 MHz [4]. It will exploit the knowledge that tracks with high transverse momentum (p_T) are often signs of interesting physics. Novel tracking modules (p_T -modules) are being developed that utilise two closely spaced silicon sensors to select only pairs of hits (*stubs*), compatible with a transverse momentum (p_T) greater than 2 GeV for level-1 triggering.

2 The Track Finding Processor

Tracks must be reconstructed from the approximately 20,000 stubs per event (at 200 pileup) within 4–5 μs if they are to be used to make the level-1 trigger decision. We propose an option for a highly time-multiplexed track-finder, using commercially available FPGAs (rather than custom ASICs), mounted on high-bandwidth processing boards. In this design, the tracker back-end consists of two layers, with the Track Finding Processor (TFP) cards separated from the Data, Trigger and Control (DTC) cards, which are responsible for the time-multiplexing of the incoming data. This design is based upon the success of the CMS Phase I calorimeter trigger [5], which utilises a similar time-multiplexed approach. The time multiplex period, T , is dictated by the data rates and available bandwidth out of the detector, and into the TFP processing board. As depicted in Fig. 1, we propose that each TFP processes only one nonant of the tracker in ϕ (motivated by a realistic tracker cabling scenario), $|\eta| < 2.4$ (where η is pseudorapidity), and one out of every 20 events. The coordinate system is described in [1].

Fig. 1. Preliminary system architecture, whereby DTCs in two neighbouring detector nonants time-multiplex and duplicate stub data across processing nonant boundaries before transmission to the TFP boards.

2.1 The Hough Transform

The Hough transform (HT) is a widely used feature extraction technique [6]. It can be used to find imperfect instances of objects within a space, for example tracks within the map of tracker hits. FPGA firmware has been developed to use a two-dimensional HT to search for primary tracks in the $r - \phi$ plane [7]. The parametrisation $(q/p_T, \phi_0)$ is used, where ϕ_0 is the coordinate of the stub at $r = 0$, q is the electronic charge, and q/p_T is the free parameter. Within the TFP, the stubs are sorted into sub-sectors in both η and ϕ . A *Hough array* is constructed for each sub-sector. Within each array, stubs are binned at each q/p_T column into a corresponding ϕ_0 row, based on the straight-line formula $\phi_0 = \phi_{\text{stub}} + 0.0015 r B q/p_T$, where B is the magnetic field strength. The

set of stubs that intercept at a point can be considered a track candidate (after passing some additional checks). An optional feature of the algorithm is the ability to use the bend of the stub between the two sensors on the p_T -modules to calculate a coarse p_T estimate, which can be used to constrain the allowed q/p_T space for each stub within the array. This *bend filter* has a configurable compatibility range, which is currently set to three standard deviations of the p_T estimate resolution. The firmware can also be configured to apply no requirement on the bend information, the consequence being a four-fold increase in the number of found track-candidates, which must be filtered out by the more precise downstream track fitting.

In firmware, each Hough array (a fully pipelined firmware block, which receives one stub per clock cycle) consists of a number of *columns*, connected to a single *book keeper*. The book keeper receives the stubs and propagates them to each column in turn. Within each column, the corresponding ϕ_0 of the stub is calculated, and a register is set to mark a hit in the relevant tracker layer in the corresponding array cell. After each time-multiplexed period, all stubs will have been recorded in the Hough array, and the tracks within cells that have been marked as having minimum required number of tracker-layers (configurable between 4 and 5 layers for each array) are read out.

2.2 The Kalman Filter

The Kalman Filter (KF) is a commonly used iterative algorithm that can turn a series of measurements containing inaccuracies and noise into estimates of unknown variables [8]. In this scenario, the measurements correspond to the stubs associated with the track-candidates produced by the HT, and the unknown variables to the final track helix parameters. A combinatorial KF has been implemented in FPGA firmware. The initial estimates (seed) of the track parameters are taken from the corresponding HT cell. The Kalman state (current estimate of the unknown parameters) is then updated, one stub at a time in increasing radius, after each cycle using a weighted average of the prior state and the new measurement. The χ^2 of the track parameters are also calculated at each stage, and this information, along with the number of layers missing hits, is used to both reject false track candidates, and remove incorrect stubs from tracks. This process is repeated until all stubs on the track are added, or a configurable time-out is reached. Following the KF, a simple Duplicate Removal (DR) algorithm removes HT candidates with duplicated helix parameters.

3 Demonstrator Hardware and Software

A hardware demonstrator has been constructed to prove the feasibility of the proposed TFP and the functionality of the firmware described above. Rather than a nonant segmentation in ϕ , the demonstrator uses an octant segmentation, as it was built based on a previous proposal. It is capable of processing the data from one octant in ϕ , $|\eta| < 2.4$, and one out of every thirty-six 40 MHz events

(equivalent to $T = 36$ events). More details on the demonstrator firmware can be found in [9].

Fig. 2. The TFP demonstrator consists of five MP7 boards, each indicated by a separate block in the diagram. Source and Sink boards represent the DTCs, and the L1 trigger, respectively. The Geometric Processor (GP) assigns stubs to the 36 sub-sectors per octant (two in ϕ and 18 in η), routing those associated with a given sub-sector to dedicated output links.

The building blocks of the hardware demonstrator are the Master Processor 7 (MP7) [10], FPGA-based, data-stream processing double-width AMC cards, equipped with a Xilinx Virtex-7 690 [11] FPGA, and 72 optical transmitters/receivers running at 10 Gbps each way. Eleven MP7 cards are installed in a MicroTCA crate at CERN. The MP7 boards are optically cabled to match the requirements shown in Fig. 2, where each MP7 board is represented by a block.

The demonstrator software package allows for the direct comparison of performance between the hardware demonstrator output and the emulator (which is based on the CMS software framework (CMSSW)), using Monte Carlo physics samples that emulate the conditions at the HL-LHC. Although the demonstrator processes one tracker octant at a time, the firmware is agnostic to the choice of octant, making it possible to take data for all octants sequentially.

4 Demonstrator Results and Performance

The hardware demonstrator has been shown to be highly efficient at track finding in simulated events containing a $t\bar{t}$ decay and 200 pileup events. Fig. 3 shows the track finding efficiency in this scenario (defined in relation to generator-level primary tracks with $p_T \geq 3$ GeV, generating stubs in at least four tracker layers within $|\eta| < 2.4$) as a function of the simulated track η and p_T , for both the emulation software and the hardware demonstrator output. The average tracking efficiency for all particles generated in the $t\bar{t}$ decay is 94.4%, and 97.1% when selecting muons from the decay. The resolution of the track parameters, as shown in Fig. 3, is approximately 1% in relative p_T and 1 mm in z_0 , in the barrel.

Fig. 3. Tracking efficiency as measured in both the hardware demonstrator and in emulation, determined using 2000 $t\bar{t}$ events each with 200 pileup events superimposed (top). Relative p_T and z_0 resolution, as measured in the emulation software, for single isolated muons, for three different ranges of p_T (bottom).

At the end of the track finding and duplicate removal chain, an average of 70 tracks per event remain. On average, 20% of these are false tracks arising from combinatorics, and 4% are duplicates of genuine tracks.

In addition to tracking performance, an important goal of the demonstrator was to prove the feasibility of track finding within the expected $4\mu\text{s}$ latency budget. Tab. 1 shows the latency for each step of the demonstrator chain. The processing latency of each stage is fixed, and is therefore independent of pileup or event occupancy.

	SerDes	GP	HT	KF	DR	Total
Latency [ns]	545	251	1025	1620	38	3479

Table 1. Processing latency for each of the demonstrator components, including the four stages of serialisation/de-serialisation (SerDes), and the optical transmission delays between each board.

5 FPGA Resources and Next Generation Hardware

The next generation of generic FPGA processing boards for CMS are in development. An example is the MPUltra PCIe card, which was developed in part to test the capabilities of the Kintex Ultrascale FPGA family [12], which support transceivers/receivers up to 16.3 Gb/s. Overall resource usage of the demonstrator TFP, as shown in Tab. 2, suggest that the functionality could feasibly fit within two Kintex Ultrascale 115 FPGA.

	LUTs [10^3]	DSPs	FFs [10^3]	BRAM (36 Kb)
GP & HT	365	3360	504	1410
KF & DR	398	5112	316	1776
Demonstrator total	763	8472	820	3186
Virtex-7 690	433	3600	866	1470
Kintex Ultrascale 115	633	5520	1266	2160

Table 2. Resource usage for each component of the TFP demonstrator, alongside available resources for some compatible Xilinx FPGAs [12].

6 Summary

A hardware demonstrator, based on currently available FPGA processing boards, has been developed to prove the feasibility of track finding for CMS at the HL-LHC. It is able to find tracks with around 95 % mean efficiency in challenging high-occupancy events (such as $t\bar{t}$ events with 200 pileup superimposed) within $4\mu s$. The results shown prove that an L1 track-finder using only FPGA processing is a feasible and safe solution for CMS. It also shows that new advances in technology are not required to achieve the desired performance, but whenever possible they will be exploited to improve system cost and latency.

References

1. CMS Collaboration, "The CMS experiment at the CERN LHC", JINST 3 (2008) S08004, doi:10.1088/1748-0221/3/08/S08004
2. THE CERN LARGE HADRON COLLIDER: ACCELERATOR AND EXPERIMENTS Collaboration, "LHC Machine", JINST 3 (2008) S08001, doi:10.1088/1748-0221/3/08/S08001
3. THE CERN LARGE HADRON COLLIDER: ACCELERATOR AND EXPERIMENTS Collaboration G. Apollinari, et al., High-Luminosity Large Hadron Collider (HL-LHC): Preliminary Design Report. CERN, Geneva, 2015. doi:10.5170/CERN-2015-005.

4. CMS Collaboration, CMS Technical Design Report for the Phase-2 Tracker Upgrade, Technical Report CERN-LHCC-2017-xxx. CMS-TDR-xxx, Geneva, month, 2017.
5. CMS collaboration Collaboration, CMS Technical Design Report for the Level-1 Trigger Upgrade, Technical Report CERN-LHCC-2013-011. CMS-TDR-12, Jun, 2013.
6. P. V. C. Hough, Method and means for recognizing complex patterns, December, 1962. US Patent 3,069,654.
7. C. Amstutz et al., An FPGA-Based Track Finder for the L1 Trigger of the CMS Experiment at the High Luminosity LHC, in 2016 IEEE-NPSS Real Time Conference (RT), pp. 19. June, 2016. doi:10.1109/RTC.2016.7543102.
8. R.Fruhworth, Application of Kalman filtering to track and vertex fitting, Nucl.Instrum.Meth. A262 (1987) 444450, doi:10.1016/0168-9002(87)90887-4.
9. R. Aggleton et al., "An FPGA Based Track Finder for the L1 Trigger of the CMS Experiment at the High Luminosity LHC", CMS-NOTE-2017-XXX; CERN-CMS-NOTE-2017-XXX- Geneva : CERN. To be published.
10. K. Compton et al., The MP7 and CTP-6: multi-hundred Gbps processing boards for calorimeter trigger upgrades at CMS, JINST 7 (2012) C12024, doi:10.1088/1748-0221/7/12/C12024.
11. Xilinx, 7 Series FPGAs Overview, Product Specification, May, 2015. DS180 (v1.17).
12. Xilinx, UltraScale Architecture and Product Data Sheet: Overview, February, 2017. DS890 (v2.11), <https://www.xilinx.com/support/documentation/data-sheets/ds890-ultrascale-overview.pdf>.